Interactive Markerless Articulated Hand Motion Tracking Using RGB and Depth Data

Srinath Sridhar, Antti Oulasvirta, Christian Theobalt
MPI Informatik and Saarland University
{ssridhar,oantti,theobalt}@mpi-inf.mpg.de
Supplementary Material, ICCV 2013

1. Training Fingertip Detection

Our fingertip detection method is based on binary classification of parts of the *depth image* into those that contain a fingertip and those that do not. We use HOG descriptors on depth images as features to train a linear SVM classifier. During classification, a sliding window is used to search for regions on the input depth image that contain fingertips. We use a window size of 32×32 for training and multi-scale detection during classification.

In order to train the linear SVM classifier, we used both real and synthetic, annotated data. Synthetic data was obtained by setting the hand model and its mesh, $\mathcal{M}$, to different poses and rendering it as a depth image. Windows of size 32×32 are extracted around fingertips on the rendered depth image. Real training data was obtained by manually annotating depth image sequences of general hand motion. Note that this data is disjoint from the standard datasets that we acquired for evaluating our method.

Table 1 shows some statistics from our training regime. We performed 4-fold cross validation and averaged the validation accuracy over all folds. We found that validation accuracy on synthetic data was higher than on real data. We therefore trained the SVM classifier using a combination of real and synthetic data and found that the validation accuracy was 91.3%. Note that the final fingertip classification accuracy is much higher since the detected fingertips are processed based on depth.

Table 1. Statistics from training phase for fingertip detection.

Training Data	**Positive Samples**	**Negative Samples**	**Accuracy (%)**
Synthetic (mesh)	540	3141	89.2
Real	867	5498	87.6
Real + Synthetic	1407	8639	91.3

2. Qualitative Results

We present more qualitative results from our method showcasing both successful tracking and failure cases. Please also see the accompanying video.

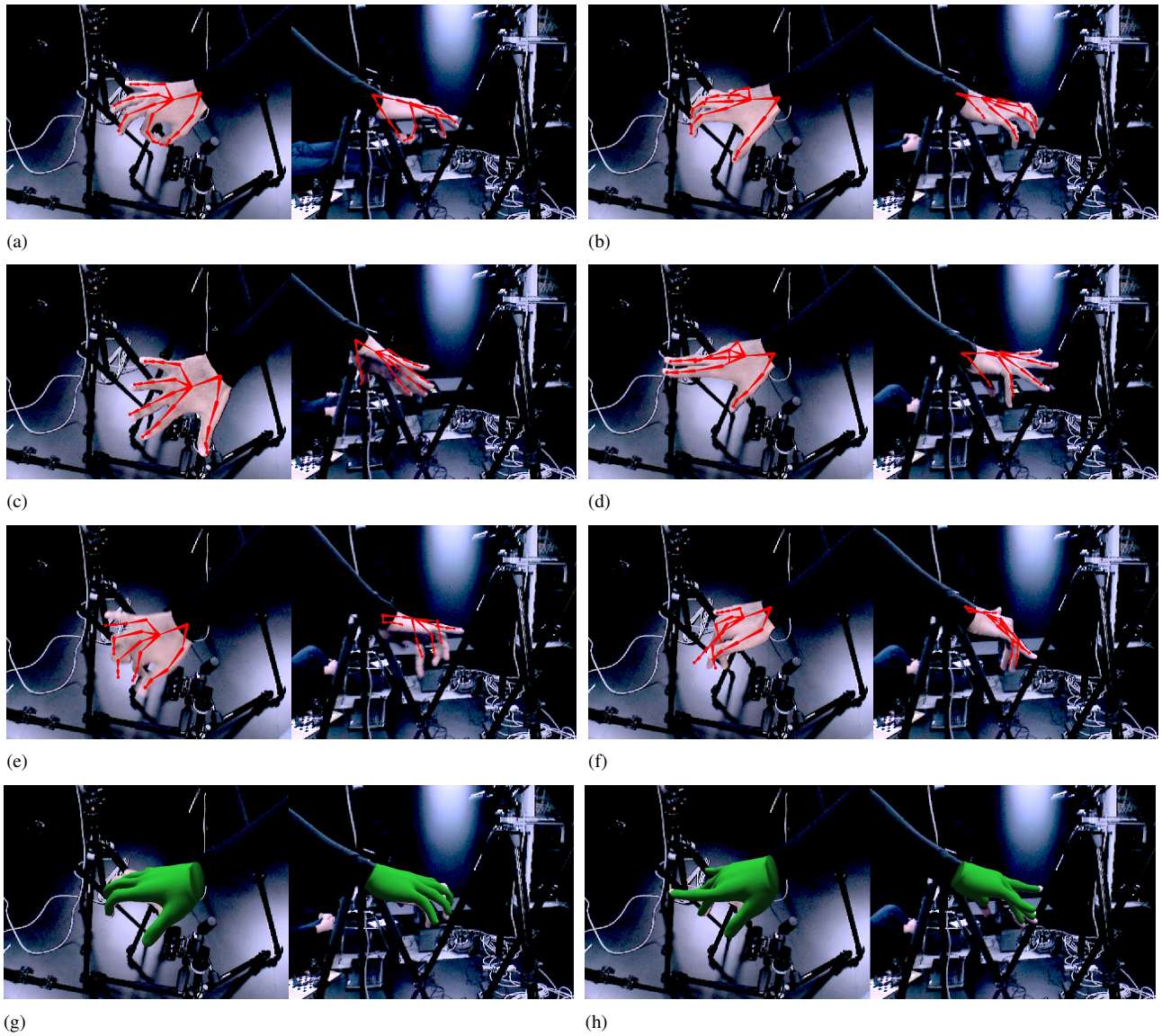

Figure 1. Qualitative results of our method as seen from two camera views. Results in (a), (b), (c) and (d) show both simple and complex motions being tracked. Result in (e) shows tracking failure due to motion blur caused by fast motion. Result in (f) shows tracking failure due to homogeneous hand colour. Results in (g) and (h) show successfully tracked frames with a mesh overlayed. Note that the mesh used does not have the exact dimensions of the hand being tracked.